

# End-to-End Automated Drum Transcription

---

## Master Research Internship Proposal

**Supervisor:** Florent Jacquemard -- florent.jacquemard@inria.fr <https://jaquema.gitlabpages.inria.fr>

**laboratory:** équipe [Vertigo](#) & Action Exploratoire Inria Codex, [Cedric, CNAM](#), 2 rue Conté 75003 Paris.

**summary:** The goal of Automatic Music Transcription (AMT) [1] is to convert the performance of a musician into music notation - somehow the analogous of speech-to-text in Natural Language Processing. It is considered to be one of the oldest and most challenging research problems in the field of Music Information Retrieval (MIR).

The case of Drum Transcription (DT) is very particular as it deals with pitchless instruments, events with (almost) no duration and specific notations. It has been the source of many MIR studies, see [2] for a survey. Most of these work focus on computational methods for the detection of drum sound events from acoustic signals, and the extraction of low-level features such as an instrument class and an onset time (peak picking). Very few however have addressed the task of the generation of readable music (rhythmic) notation from the above features, a step crucial in a musical context and far from being trivial.

This proposal is concerned with the later problem, and more precisely, with:

1. the study of **Language Models** (LM) embedding some high-level musical information needed for the generation of quality music score. One should in particular consider hierarchies of drum events inducing consistent temporal placements amenable to rhythm notations easy to read for a trained drummer; see [3] for related tree structured models based on formal language theory, that we are developing in the context of more general purpose AMT tools.
2. the **integration** of these LM with state of the art **Acoustic Models** (AM) and methods for the above signal processing tasks. That requires taking into account the musical context and high level

musical information of LM on top of the above low level acoustic features.

Furthermore, some experiments will be conducted on the base of public datasets, in order to evaluate the integrated approach. They should cover some cases of complex overlapping rhythmic patterns.

Beyond model integration, it will also be interesting to study how the use of LM can improve the results of AM see [2], and pave the way to fully automated drum score generation and the general problem of end-to-end AMT.

## References

- [1] Automatic music transcription: challenges and future directions. E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri. Journal of Intelligent Information Systems, vol. 41, no. 3, pp. 407–434, 2013.
- [2] A Review of Automatic Drum Transcription. C.-W. Wu et al. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(9) 2017.
- [3] A Parse-based Framework for Coupled Rhythm Quantization and Score Structuring. F. Foscarin, F. Jacquemard, P. Rigaux, and M. Sakai. 7th International Conference on Mathematics and Computation in Music (MCM 2019). see also <https://qparse.gitlabpages.inria.fr>